Ruiz-Sanchez AJ, Higgs PL, Peters DT, Turley AT, Dobson MA, North AJ, Fulton DA.

Probing the Surfaces of Biomacromolecules with Polymer-Scaffolded Dynamic Combinatorial Libraries.

***ACS Macro Letters* 2017, 6(9), 903-907.**

Probing the surfaces of biomacromolecules with polymer-scaffolded dynamic combinatorial libraries

Antonio J. Ruiz-Sanchez,[a] Patrick L. Higgs,[a] Daniel T. Peters,[b] Andrew T. Turley,[a] Matthew A. Dobson,[a] Adam J. North[a] and David A. Fulton*[a]

a Chemical Nanoscience Laboratory, School of Chemistry, Bedson Building, Newcastle University, Newcastle upon Tyne, NE1 7RU, UK.

b Institute for Cell and Molecular Biosciences, Newcastle University, Framlington Place, Newcastle upon Tyne NE2 4HH, United Kingdom

ABSTRACT: Methods to analyze and compare biomacromolecular surfaces are still in their relative infancy on account of the challenges involved in comparing surfaces computationally. We describe a systems chemistry approach that utilizes polymer-scaffolded dynamic combinatorial libraries to experimentally probe biomacromolecular surfaces in aqueous solution which provides feedback as to the nature of the surfaces, allowing the comparison of three globular proteins and a nucleic acid.

The surfaces of biomacromolecules are determined by their shapes and physicochemical properties—including their electrostatic potentials, hydrophobicities and hydrophilicities—which influences their function and interactions with other species.[1] Although the structures of many biomacromolecules are known (the Protein Data Bank contains the 3-D structures of over 127,000 proteins and nucleic acids), methods to analyse and compare biomacromolecular surfaces are still in their relative infancy on account of the challenges involved in comparing surfaces computationally.[2] Furthermore, surface hydration—another key surface parameter—is computationally very difficult to model, further adding complexity to the problem.[3]

We reasoned that an approach to experimentally interrogate the surfaces of biomacromolecules in aqueous solution could be a potentially useful tool, providing feedback as to the nature of biomacromolecular surfaces and allow relative comparisons between biomacromolecules to be drawn. We have applied a systems chemistry[4] approach in which we employ so-called polymer-scaffolded dynamic combinatorial libraries (PS-DCLs)[5-10] (Fig 1), which we have shown respond to the addition of macromolecular templates by changing their compositions, preferentially incorporating residues which promote binding and rejecting residues which do not. We hypothesized that these compositional changes are likely to reflect the surface properties of the biomacromolecular template. The response of a single PS-DCL provides limited information, but we envisioned that collating the responses of a series of different PS-DCLs into a response 'profile' would provide a more comprehensive description. Analysis of the response profiles would then allow an assessment of the relative similarities of the biomacromolecular templates. Otto and coworkers have used[11] DCLs of disulfide macrocycles to assess the similarities of small molecules, showing that analysis of library responses allowed differences between ethylamine-containing molecules and other amines/ammoniums to be distinguished, demonstrating the potential of DCLs to discriminate molecular surfaces. Here we demonstrate that PS-DCLs can be harnessed to perform a rudimentary comparison of the surfaces of biomacromolecules, confirming the similarity of two near-identical proteins and distinguishing between globular proteins and a nucleic acids.

PS-DCLs were constructed using the aldehyde-functionalised poly(acrylamide) scaffold **P1** (Fig. 2), and the four acylhydrazide residues **R1**-**R4** were chosen to display functionalities which may interact with macromolecular templates through ion-ion (**R2**/**R3**), ion-dipole, or dipole-dipole interactions (**R1**/**R4**). This arrangement makes these PS-DCLs particularly sensitive to the electrostatic surface potentials of biomacromolecules, and we anticipate that the palette of residues can in principle be expanded to also allow the hydrophobicities of surfaces to be probed. Two structurally similar proteins—bovine serum albumin (BSA) and human serum albumin (HSA)— as well as the protein pepsin and a double-stranded DNA (from calf thymus) were selected as biomacromolecular templates.

A 4-residue PS-DCL was prepared (see ESI) from polymer **P1** and residues **R1**-**R4** in buffered D_2O (NH_4OAc/AcOH pD 4.5). The BSA template, which under the conditions of our experiments displays a net positive charge (pI 5.5), was added and after 24 h ^{1}H NMR spectroscopic integral analysis revealed re-equilibration was completed. We have demonstrated previously[6-9] that the PS-DCL residue compositional changes upon template addition can be monitored indirectly by ^{1}H NMR spectroscopy. We have continued to utilize this approach as the composition of the residues conjugated to the polymer scaffolds cannot be monitored directly because diagnostic signals corresponding to conjugated residues are broad, making their accurate integration difficult. Thus, ^{1}H NMR spectroscopy was used to measure the relative concentrations of unconjugated residues **R1**-**R4** in solution, allowing the residual composition upon the polymer scaffolds to be ascertained. Integral analysis of the α-methylene resonances of **R1**-**R4** was used to determine the mole fractions of the unconjugated residues in solution before and after templation (see

ESI). Compositional changes in response to biomacromolecular template addition are best shown in the graphical form of a results 'profile' (Fig 3a), where residues preferentially incorporated onto the polymer scaffolds display negative values and residues rejected by the polymer scaffolds display positive values of change in mole fraction. The experimentally determined limits of quantification (see ESI, Fig. S1 and S2) are also shown in this figure, and changes in mole fraction which fall below these limits were considered as being too small to quantify with confidence. The result indicates the 4-residue PS-DCL has responded to addition of the biomacromolecular template, BSA, to incorporate a greater proportion of anionic residue **R3,** rejecting the other residues. We presume residue **R3** was preferentially incorporated on account of favorable interactions between anionic residues upon the polymer and cationic patches upon the template.

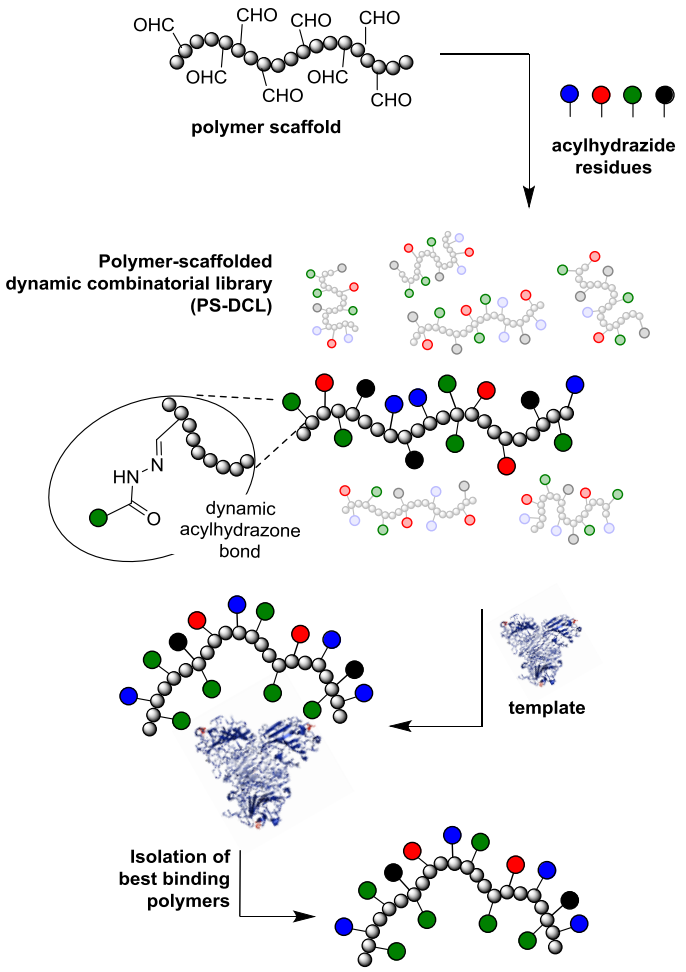


Fig. 1. Conjugation of acylhydrazide residues onto an aldehyde-functionalised polymer scaffold generates a polymer-scaffolded dynamic combinatorial library of polymers which can interconvert by exchanging residues. Each residue type features different functionality. The library responds to the addition of template by undergoing compositional change,[6,7] preferentially incorporating residues which promote binding to the template.[8,9] It is our hypothesis that these compositional changes are likely to reflect the surface properties of the biomacromolecular template.

To gain a more comprehensive description of PS-DCL response to biomacromolecular template addition we then prepared and templated all possible permutations of three residues

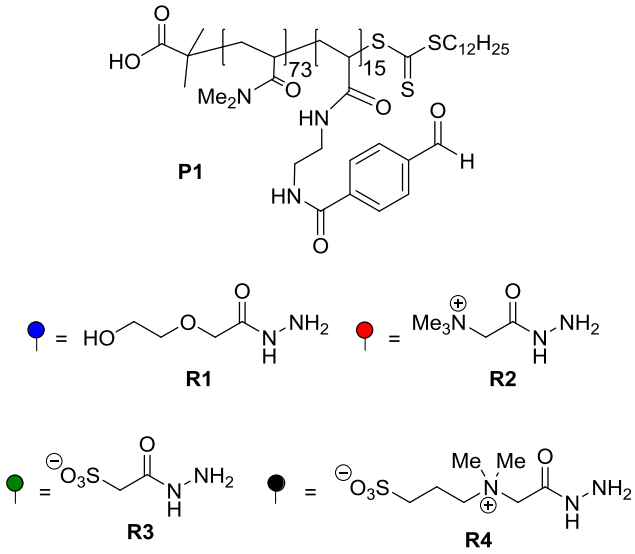


Fig. 2 Structure of polymer scaffold **P1** and residues **R1**-**R4**.

(4 x 3-residue libraries) and two residues (6 x 2-residue libraries). The systematic removal of one then two residues from the original library constitutes a dynamic deconvolution, an approach to library analysis which has been applied previously[12-16] by researchers to identify high affinity building blocks within DCLs. The responses of all 4-, 3- and 2-residue PS-DCLs (11 libraries in total) were then collated (Fig. 3a). The responses of the 3-residue libraries to BSA template addition were dominated by preferred incorporation of anionic **R3** at the expense of the other residues. The **R1**/**R2**/**R4** library, which is absent in the anionic **R3**, displays little compositional change, an observation that suggests that these residues are neither particularly favored nor disfavored and thus do not interact strongly with the template. The responses of the 2-residue libraries to template addition (Fig 3a) provides insight to this observation. The libraries **R1**/**R4** and **R2**/**R4** appear to undergo little compositional change when templated, further supporting the idea that these residues have little interaction with the surface of the biomacromolecular template and thus their incorporation is neither favored nor disfavored. The response of library **R1**/**R2**, however, indicates that the cationic residue **R2** is rejected and the neutral residue **R1** is incorporated,[17] an observation which is contrary to the idea that **R1** and **R2** are neither particularly favored or disfavored. The fact that the removal of **R4** from the **R1**/**R2**/**R4** library leads to unexpected behavior in the **R1**/**R2** library highlights the usefulness of library deconvolution to dissect the nature of library response to biomacromolecular surfaces. The responses of the 2-residue libraries featuring the anionic residue **R3** were as anticipated, with the libraries responding by favorably incorporating **R3** at the expense of the other residues. Taken together, the response 'profile' suggests that library responses to BSA template addition were dominated by the preferred incorporation of anionic residue **R3**, and that this residue is most likely key in the stabilization of polymer-template complexes. The results also suggest that on the balance of evidence obtained, the other residues are not particularly favored nor disfavored by this template.

The results of templating experiments with HSA are shown in Fig 3b. The response 'profile' displays very similar

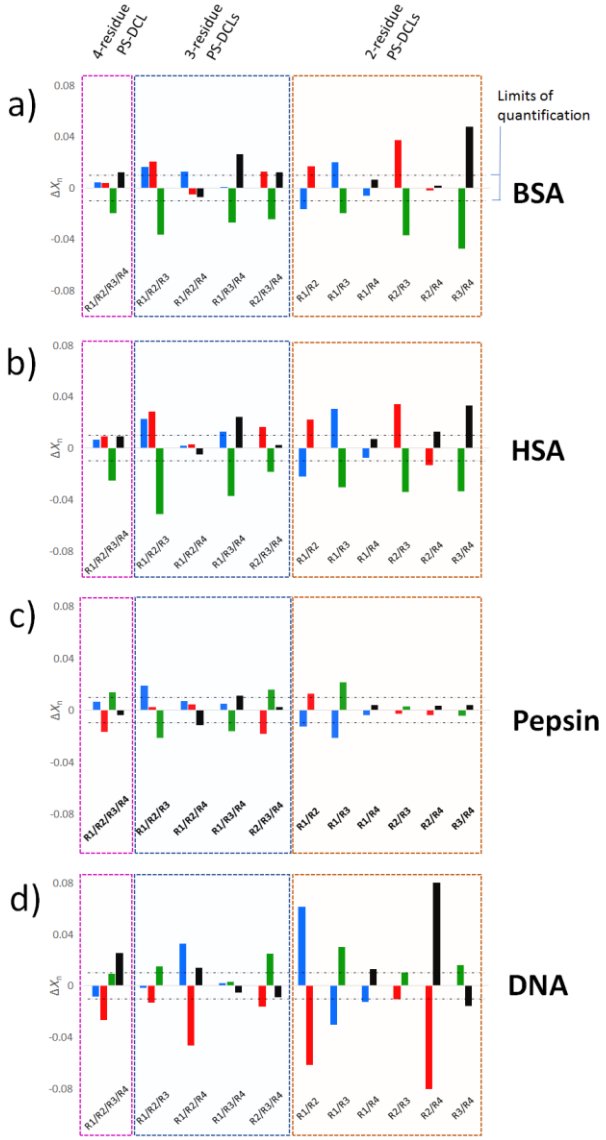


Fig 3. Compositional changes experienced by the 4-residue, the four 3-residue and six 2-residue PS-DCLs upon exposure to templates (a) BSA, (b) HSA, (c) pepsin and (d) DNA. Compositional change is presented as change in mole fraction of each residue (ΔX_n) within the library. The limits of quantification (see ESI) are represented by dashed lines.

trends to that observed for the structurally similar BSA, an observation suggesting a correlation exists between PS-DCL response to templation and the surface of the template. To assess the degree of similarity between the surfaces of HSA (PDB code: 4K2C) and BSA (PDB code: 4F5S), the 3D-SURFER tool[18] was used, revealing a Euclidean distance of 1.650 between the 3D Zernike descriptors that describe the surface of each protein. Distances of less than 10 are indistinguishable from an identical protein that is in a different orientation.[19] A global analysis of structural similarity was also performed by superposition of the two structures (Fig. S3) using the GESAMT superpose routine of the CCP4MG software package.[20,21] The superposition returned a root mean squared deviation (RMSD) of 1.8 Å between the two structures over 569 Cα atoms. This result corresponds with a previous comparison of the structures of HSA and BSA, which reported an RMSD of 1.5 Å between the two proteins.[22] These analyses show that both the structure and surface of the two proteins are highly similar, which correlates well with the high similarity of the response 'profiles' obtained in the templating experiments.

To investigate the basis of the templation experiments further, electrostatic surface representations of both HSA and BSA were generated using CCP4MG (Fig. S4). Analysis of these surfaces revealed that in both structures the charge is not uniformly distributed. Instead, the structures contain both a large positively charged cleft, as well as a patch of clustered negative charge on the opposite face of the protein. Although the actual surface charges present on the proteins in the templating experiment will differ from these representations (as the experiments were not conducted at physiological pH), the presence of these charged regions may go some way towards explaining how anionic residue **R3** and cationic **R2** could interact with the proteins. In particular, these observations provide insights as to why **R2** is not particularly disfavored; it may be that by interacting with the negative patches of surface charge, or simply being located in areas which avoid interaction with cationic patches, **R2** can happily be accommodated within the polymer-template complex.

We then templated our PS-DCLs with pepsin, whose structure is very different to BSA/HSA (Figure S3). Pepsin possesses a low pI value of 2.2 [23] on account of the significant display of acidic residues upon the protein surface and aspartic acid residues present within special microenvironments that influence their pK_a values.[24] Consequently, pepsin displays a very different electrostatic surface representation to BSA/HSA (Fig. S4). The response profile of the collated PS-DCLs towards pepsin (Fig. 3c) is very different to those observed for BSA/HSA, a significant observation which lends support to the idea that PS-DCLs are able to differentiate between bioacromolecular surfaces. The response of the 4-residue PS-DCL (Fig 3c) shows that the cationic residue **R2** is preferentially incorporated, presumably on account of favorable ion-ion interactions between cationic charges on the polymer scaffold and the anionic nature of the protein surface. The response of the 3-residue library **R2**/**R3**/**R4** was as anticipated, with incorporation of **R2** at the expense of other residues suggesting these polymers are also interacting with regions of the surface that display anionic charge. The response of the **R1**/**R3**/**R4** library shows the favorable incorporation of the anionic residue **R3**, an unexpected observation which at this time we cannot fully rationalize, but suggests the interactions of polymer is templates is more complex than 'simple' electrostatics'. We speculate the same rationale explains the response of the **R1**/**R2**/**R3** library and the 2-residue libraries **R1**/**R2** where the cationic residues are rejected.

We then templated our PS-DCLs with DNA (from calf thymus) (Fig 3d). The response of the 4-residue PS-DCL shows that cationic residue **R2** is preferentially incorporated, presumably on account of favorable ion-ion interactions between cationic residues upon the polymer and the anionic phosphodiesters of the template. The responses of the 3-residue libraries indicated that **R2** was also preferentially incorporated, largely at the expense of other residues, further supporting the idea that this residue is involved in favorable interactions with the surface of the template. The negligible response of the **R1**/**R3**/**R4** library suggests these three residues are neither favored nor disfavored, and is an observation strongly reminiscent of the response of the **R1**/**R2**/**R4** library with BSA/HSA. Again, the responses of the 2-residue PS-DCLs **R1**/**R3**, **R1**/**R4** and **R3**/**R4**—which do

not include the favorable **R2** residue—are intriguing. In particular, a strong response was observed in the **R1**/**R3** library, suggesting libraries featuring **R1** and **R3** respond differently when **R4** is not present. Taken together the results of these three templating experiments further suggests there can be a complex interplay between polymer and template which is not always straightforward to rationalize. Strong responses observed with PS-DCLs **R1**/**R2** and **R2**/**R4** indicate cationic residue **R2** was preferentially incorporated at the expense of **R1** or **R4**, further supporting the involvement of **R2** in favorable electrostatic interactions with the polyanionic template. Importantly, the response 'profile' obtained with the DNA template is very different to those obtained with the globular proteins BSA, HSA and pepsin, an observation which suggests that through their interactions with the surfaces of biomacromolecular templates PS-DCLs are able to form a rudimentary comparison of their surfaces.

In conclusion, collating together the responses of PS-DCLs to biomacromolecular template addition affords response profiles that enable comparisons. Similarities in response 'profiles' were observed between two structurally similar proteins (BSA and HSA), and these were markedly different to the 'profile' of pepsin and a DNA template, whose surfaces are very different in nature to those of BSA/HSA. These observations suggest that PS-DCL response correlates with biomacromolecular surface and that PS-DCLs are a promising tool in probing the surfaces of biomacromolecules. To explore this idea further we are exploring PS-DCLs prepared utilizing a wider palette of residue types—including hydrophobic residues which we anticipate will enable the probing of surface hydrophobicity—and investigating their responses to a wider selection of templates. We also believe that intramolecular crosslinking of the polymer scaffold will provide improved rigidity and reduce entropic penalties when polymers interact with biomacromolecular surfaces, leading to more differentiated response profiles.

Supporting Information. Synthetic procedures for the synthesis of **P1**, **R1**, **R3**-**R4**, preparation of PS-DCLs, determination of limits of quantification and structural comparisons of BSA with HSA are included. This material is available free of charge via the Internet at http://pubs.acs.org.

Corresponding Author

*Email: david.fulton@ncl.ac.uk.

ACKNOWLEDGMENT

We thank Dr F. J. Lowes for help with the treatment of experimental errors.

REFERENCES

1. Burgoyne, N. J.; Jackson, R. M. "Predicting protein function from surface properties". In: Rigden, DJ., editor. From Protein Structure to Function with Bioinformatics. Springer; Dordrecht, Netherlands: 2009. p167-186.

2. Yang, H.; Qureshi R.; Sacan, A. Protein surface representation and analysis by dimension reduction. *Proteome Science* **2012**, *10* (Suppl 1):S1.

3. Jeszenői, N.; Horváth, I.; Bálint, M.; van der Spoel D.; Hetényl, C. Mobility-based prediction of hydration structures of protein surfaces. *Bioinformatics* **2015**, *31*, 1959-1965.

4. Ashkenasy, G.; Hermans, T. M.; Otto S.; Taylor, A. F. Systems chemistry. *Chem. Soc. Rev*., **2017**, *46*, 2543-2554.

5. Fulton, D. A. Dynamic Covalent Combinatorial Libraries Constructed on Polymer Scaffolds. *Org. Lett*., **2008**, *10*, 3291-3294.

6. Mahon, C. S.; Jackson, A. W.; Murray B. S.; Fulton, D. A. Templating a polymer-scaffolded dynamic combinatorial library. *Chem. Commun*., **2011**, *47*, 7209-7211.

7. Mahon, C. S.; Jackson, A. W.; Murray B. S.; Fulton, D. A. Investigating templating within Polymer-Scaffolded Dynamic Combinatorial Libraries. *Polym. Chem*., **2013**, *4*, 368-377.

8. Mahon C. S.; Fulton, D. A. Templation-induced re-equilibration in polymerscaffolded dynamic combinatorial libraries leads to enhancements in binding affinities. *Chem. Sci*., **2013**, *4*, 3661-3666.

9. Mahon, C. S.; Fascione, M. A.; Sakonsinsiri, C.; McAllister, T. E.; Turnbull, W. B.; Fulton, D. A. Templating Carbohydrate-Functionalised Polymer-Scaffolded Dynamic Combinatorial Libraries with Lectins. *Org. Biomol. Chem*., **2015**, *13*, 2756–2761.

10. Mahon C. S.; Fulton, D. A. Mimicking Nature with Synthetic Macromolecules Capable of Recognition. *Nat. Chem*., **2014**, *6*, 665-672.

11. Saggiomo, V.; Hristova, Y. R.; Ludlow R. F.; Otto, S. Systems chemistry: using thermodynamically controlled networks to assess molecular similarity. *J. Syst. Chem*. **2013**, 4:2.

12. Bunyapaiboonsri, T.; Ramström, O.; Lohmann, S.; Lehn, J.-M.; Peng L.; Goeldner, M. Dynamic Deconvolution of a Pre-Equilibrated Dynamic Combinatorial Library of Acetylcholinesterase Inhibitors. *ChemBioChem*, **2001**, *2*, 438-444.

13. Ramström, O.; Lohmann, S.; Bunyapaiboonsri T.; Lehn, J.-M. Dynamic Combinatorial Carbohydrate Libraries: Probing the Binding Site of the Concanavalin A Lectin. *Chem. Eur. J*., **2004**, *10*, 1711-1715.

14. Abelrahim, M. Y. M.; Tanc, M.; Winum, J.-Y.; Supuran C. T.; Barboiu, M. Dominant behaviours in the expression of human carbonic anhydrase hCA I activity. *Chem. Commun*., **2014**, *50*, 8043-8046.

15. Schaufelberger F.; Ramström, O. Dynamic Covalent Organocatalysts Discovered from Catalytic Systems through Rapid Deconvolution Screening. *Chem. Eur. J*., **2015**, *21*, 12735-12740.

16. Jiang, Q.-Q.; Sicking, W.; Ehlers M.; Schmuck, C. Discovery of potent inhibitors of human *β*-tryptase from pre-equilibrated dynamic combinatorial libraries. *Chem. Sci*., **2015**, *4*, 1792–1800.

17. The response of the **R1**/**R2** PS-DCL may arise on account of favourable ion-dipole interactions between **R1** and the template, or in avoidance of unfavourable cation-cation interactions between **R2** and the template, or a combination of both factors.

18. La, D.; Esquivel-Rodriguez, J.; Venkatraman, V.; Li, B.; Sael, L.; Ueng, Ahrendt, S.; Kihara, D. 3D-SURFER: software for high-throughput protein surface comparison and analysis. *Bioinformatics* **2009**, *25*, 2843-2844.

19. Sael, L.; Li, B.; La, D.; Fang, Y.; Ramani, K.; Rustamov R.; Kihara, D. Fast protein tertiary structure retrieval based on global surface shape similarity. *Proteins* **2008**, *72*, 1259-1273.

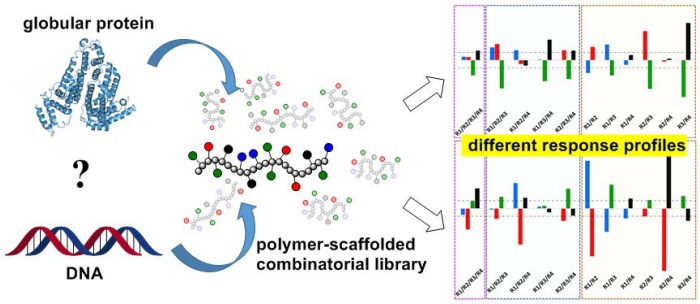
20. Krissinel, E. Enhanced fold recognition using efficient short fragment clustering. *J. Mol. Biochem*., **2012**, *1*, 76-85.

21. McNicholas, S.; Potterton, E.; Wilson K. S.; Noble, M. E. M. Presenting your structures: the *CCP*4*mg* molecular-graphics software. *Acta Crystallogr., Sect. D: Biol. Crystallogr*., **2011**, *67*, 386-394.

22. Bujacz, A. Structures of bovine, equine and leporine serum albumin. *Acta Crystallogr., Sect. D: Biol. Crystallogr*., **2012**, *68*, 1278-1289.

23. Malamud, D., Drysdale, J. W. Isoelectric points of proteins: A table. *Anal. Biochem.*, **1978**, *86*, 620-647.

24. Andreeva, N. S.; James, M. N. G. B.M. Dunn, editor: Structure and Function of the Aspartic Proteinases. Plenum Press, New York, 1991. p 39-45



Insert Table of Contents artwork here